ELSEVIER

PII: S0308-8146(96)00112-4

Food Chemistry, Vol. 57, No. 1, pp. 127–131, 1996

0308-8146/96 $15.00+0.00

Development of food composition database management systems: the New Zealand experience

Barbara Burlingame

New Zealand Institute for Crop and Food Research, Private Bag 11030, Palmerston North, New Zealand

New Zealand has a dynamic, continually-evolving, nutritionally-appropriate and technologically-relevant food composition data system due to two main factors: the core team has always been multidisciplinary and INFOODS recommendations have been followed and implemented. From its inception, the food data team included analytical chemists and biochemists for data generation, nutrition scientists for data compilation and dissemination, and programmers and systems analysts to keep the database current and to exploit the information technology resources as they were developing. INFOODS can be credited with developing systems and strategies critically important in New Zealand, and addressing many of the major problems in data compilation. Some features and uses of New Zealand's data system are described, including implementation of tagnames, raw data capture, food images, project management from sampling plan to user products and experiences with interchanging data files. Copyright © 1996 Elsevier Science Ltd

INTRODUCTION

New Zealand's food composition database development evolved as a logical extension of its work in the area of feed composition databases, which began in the 1970s in concert with INFIC's (International Feed Information Centres) efforts, when possible. The demand for high-quality food composition data was exceeding the demand for data on feed so, in 1986, New Zealand's food composition database was established as a separate system from the feed database (Shields *et al.*, 1988). This coincided with our awareness of another group, INFOODS (the International Network of Food Data Systems), which had been established a few years after INFIC, under the auspices of United Nations University (UNU).

Fundamental to New Zealand's ability to keep its food database current and to exploit the information technology resources as they were developing, was the multidisciplinary team involved in the nutrition programme. From its inception, the food data team included analytical chemists and biochemists for data generation, computer personnel as programmers and systems analysts for data compilation and dissemination, and nutrition scientists involved in all three activities. Also fundamental was the realization that a group like this gains advantages by working with others in the field. Paying attention to INFOODS as it developed was critically important in New Zealand because it was addressing the major problems in data compilation that had not been addressed before.

INFOODS' IMPACT

Although data generators had a long-standing system for worldwide collaborations, interlaboratory proficiency trials and many refereed journals, data compilers did not have an international forum for their research. In 1987, the UNU/INFOODS established the *Journal of Food Composition and Analysis*, published by Academic Press and edited by Dr Kent Stewart. The journal, now in its eighth year, covers all scientific aspects of the data on chemical composition of human foods with particular emphasis on analytical methods for obtaining data; actual data on composition of foods; and studies on the manipulation, identification, statistics, storage, distribution and use of food composition data. This refereed journal has contributed to the greater awareness of food composition activities as legitimate research, enabling scientists to pursue this field in a more spirited fashion.

One fundamental issue INFOODS has tackled is the unambiguous identification of nutrients and other food components. The book *Identification of Food Components for INFOODS Data Interchange* (Klensin *et al.*, 1989), and its frequently updated counterpart on the World Wide Web,[a] lists international nutrient identifiers, also known as tagnames. The use of tagnames allows proper interpretation of what is meant by common nutrient names, such as *carbohydrates* (e.g. total by difference vs available by difference vs available by summation vs available by summation in monosaccharide equivalents) and *protein* (e.g. total calculated from amino nitrogen vs total by direct analysis vs total calculated from total nitrogen). Tagnames encompass the component entity, a default unit of measure, the methods of analysis where different methods produce different results and, in some cases, require 'keywords' for factors used in some component calculations (e.g. nitrogen conversion factors with the protein tag).

New Zealand was the first nation in the world to incorporate tagnames into their food composition database in 1991 (Burlingame, 1991*a*,*b*, 1993). Currently, many databases use tagnames, including the Pacific Islands Food Composition Data Base, and databases in ASEANFOODS and LATINFOODS Region. In the up-coming revision to the Australian and USDA database, tagnames will be included (Lewis, 1995; Haytowitz, personal communication).

Use of tagnames is the first essential step in engaging in international interchange of food composition data. Another book, *INFOODS Food Composition Data Interchange Handbook* (Klensin, 1992), presents the theoretical details on food composition data structure and rules for moving data files between countries and regional organizations in a way that preserves all the available information. Interchange of food composition data files has taken place between the co-operating organizations in OCEANIAFOODS, and inter-regionally.

Current issues being examined by INFOODS' committees in the hopes of attaining some standardization include data quality indications, food nomenclature and terminology, and the next generation of interchange formats.

NEW ZEALAND'S NUTRITIONAL INFORMATION SYSTEMS

The Nutritional Information Systems Data Base was originally developed to run on a Novell Network in an MS-DOS environment, using Advanced Revelation (3.1) as the development platform. This system has been described in many papers (Burlingame *et al.*, 1990; Burlingame, 1991*a*,*b*; Milligan & Burlingame, 1991). A new version in a Windows environment has been developed using Paradox, and a newer system in a client–server platform is in the works. The program is the tool used from the sample planning stage through the data dissemination stage. Table 4 shows an example of menus.

[a]URL: http://www.crop.cri.nz/crop/infoods/infoods.html

SELECTED FEATURES

The new system has been designed to interact with many other software products as modules, and to launch other applications and information resources, rather than trying to re-create existing specialty products. Some of the modules include handbooks of methods, handbooks of food additives and ingredients, business directories (for food sector), dictionaries of foods and images of foods, all on disk or CD-ROM. Also programmed for interaction are project management software, publishing software and the interface standard developed by the United States Food and Drug Administration (Petersen, 1994).

Presently, this Information Systems Data Base contains nine separate databases of various sizes and for various purposes (Table 1).

Foods are described or named in multiple facets (Truswell *et al.*, 1991). In our investigations of the optimal 'naming' system for foods, it was clear that there was some overlap between what is typically considered sample documentation, food name and food descriptors. A file for sample documentation was created to include information on sample collection and handling (Greenfield & Southgate, 1992). Food descriptors in facets and food name were considered to be one and the same. Table 2 shows the 'name' or 'descriptor' facets. At present there are 11 facets that constitute a FULLNAME, with only Generic requiring an entry (i.e. 2–11 are optional). Facet 12, SHORTNAME, is a combination of the other facets, with a 32-character limit that sometimes dictates abbreviations and selection of only enough information to make the Shortname unique. This field was created to accommodate data products such as applications software packages and concise

Table 1. Databases within the New Zealand Nutritional Information Systems

Database	Size (Mbytes)	Purpose
NZ Food Composition[1]	38.9	Maintenance; development; research; output
NZ Recipes[1]	10.9	Development; research
SPC[2]	6.0	Development assistance; interchange
INCAP[3]	1.7	Development assistance; interchange
USDA[4]	37.2	Ftp pickup; interchange
UK[5]	0.07	Demonstration copy
Thai: MOPH[6]	0.2	Interchange
Thai: INMU[7]	0.3	Development assistance; interchange
Australia[8]	6.0	Interchange

[1]New Zealand Institute for Crop and Food Research, Palmerston North. [2]South Pacific Commission, Noumea, New Caledonia. [3]Institute of Nutrition for Central America and Panama, Guatemala City, Guatemala. [4]US Department of Agriculture. [5]Royal Society of Chemistry, UK. [6]Ministry of Public Health, Bangkok, Thailand. [7]Institute of Nutrition Mahidol University, Salaya, Nakhon Pathom, Thailand. [8]National Food Authority, Canberra, Australia.

Table 2. Multifaceted naming system

1.	Generic
2.	Kind
3.	Strain
4.	Part
5.	Process
6.	Grade
7.	Maturity
8.	Genus
9.	Species
10.	Variety
11.	Message
12.	Shortname
13.	Alternative name(s)
14.	Langual codes

printed tables where space is a consideration. Shortname is the only other facet besides Generic that requires an entry. The INCAP name facets are in Spanish, and the MOPH and INMU names implement the Thai character set in the alternative name facet. Standards for food nomenclature and terminology are being further developed by an expert committee, convened by IUNS and INFOODS. New Zealand will be involved in the process and will implement the recommendations once there is agreement.

In New Zealand we were able to work through problems as they arose, and deal with complex data compilation situations systematically and on an *ad hoc* basis. Some of the features of New Zealand's system implemented from the start include the provision for input of dry matter values, and recalculation to a wet weight basis with an appropriate water value; standard error calculations that accommodate the standard deviation of the water value and the nutrient in question; source/quality codes to identify the origin and quality of the data; auto-calculation features with overrides, such as would be required for energy from a fat value that included wax esters; and a suite of integrity tests that monitors and reports on the relationships between different nutrients in the same food, and the same nutrients in similar foods.

From the early 1990s other features have been put into place. One important feature is the recipe calculation program. There are two major uses for this program: the obvious one for calculation of components of recipes with nutrient retention factors for different cooking methods; and also for preparation of records requiring weighting of proportions, such as different separable lean and separable fat ratios for meats or combinations of different cultivars of fruit based on market share data to calculate a 'combined cultivar' record. This program relies on standard equations (Rand *et al.*, 1991), nutrient retention factors (USDA, 1984 and onward) as a pop-up; as well as customized equations and retention factors and user-selected yields.

The raw data program allows capture of individual sample results when multiple samples have been collected, and individual replicate determinations on a single sample. All values can be captured and individual values, such as suspicious ones, can be excluded from further processing. Analytical data can be entered on a wet weight basis, a dry matter basis or on a freeze-dried basis, depending on how the samples were prepared and analysed. The program calculates sample means from the replicates, then calculates the means of the samples and processes the statistical information to present modes, medians, means, ranges, and standard deviations and standard errors. Three different outlier subroutines are built into the system, to be used when justified by the number of individual measurements. Trace data, captured as the limit of detection or limit of quantitation, can be processed in any of four ways, as the compiler deems appropriate: as zero; as one-half the value; as the value itself; or simply as TRACE. An analytical methods database is called by this program, with pop-ups for selection of the method of analysis used for each nutrient. The feature has been implemented in the replicate data segment because the same sample can be analysed by more than one method, for example, selenium by ICP first and again by fluorometry.

A CHANGES program audits the data over time. A *reason* field in this program requires an explanation code whenever a modification in a single or mean value is made. This will allow distinctions to be made in the database between a corrected error; an updated value which is more representative than the older one; and a

Table 3. Information provided in the different output forms from the New Zealand Food Composition Data Base

Output form	Foods	Components	Basis	Numeric data	Source/quality codes
FOODfiles	All (~1650)	52–423, according to users' needs	Per 100 g e.p.; amino acids in mg/gN, fatty acids in g/100 g TFA	Mean, standard deviation, standard error, number of samples	Complete for each nutrient in each record
Diet1/NZ	All	Subset of 52	Per 100 g or any serving size as user selection	Mean	Not provided
Tables, unabridged	All	All	Per 100 g e.p.; amino acids in mg/gN, fatty acids in g/100 g TFA	Mean, standard deviation, standard error, number of samples	Complete for each nutrient in each record
Tables, abridged	All	Subset of 52	Per 100 g and one common serve	Mean	Complete for each nutrient in each record
Tables, concise	Subset of ~800	Subset of 28	Per 100 g and up to two common serves	Mean	Single majority source listed in index

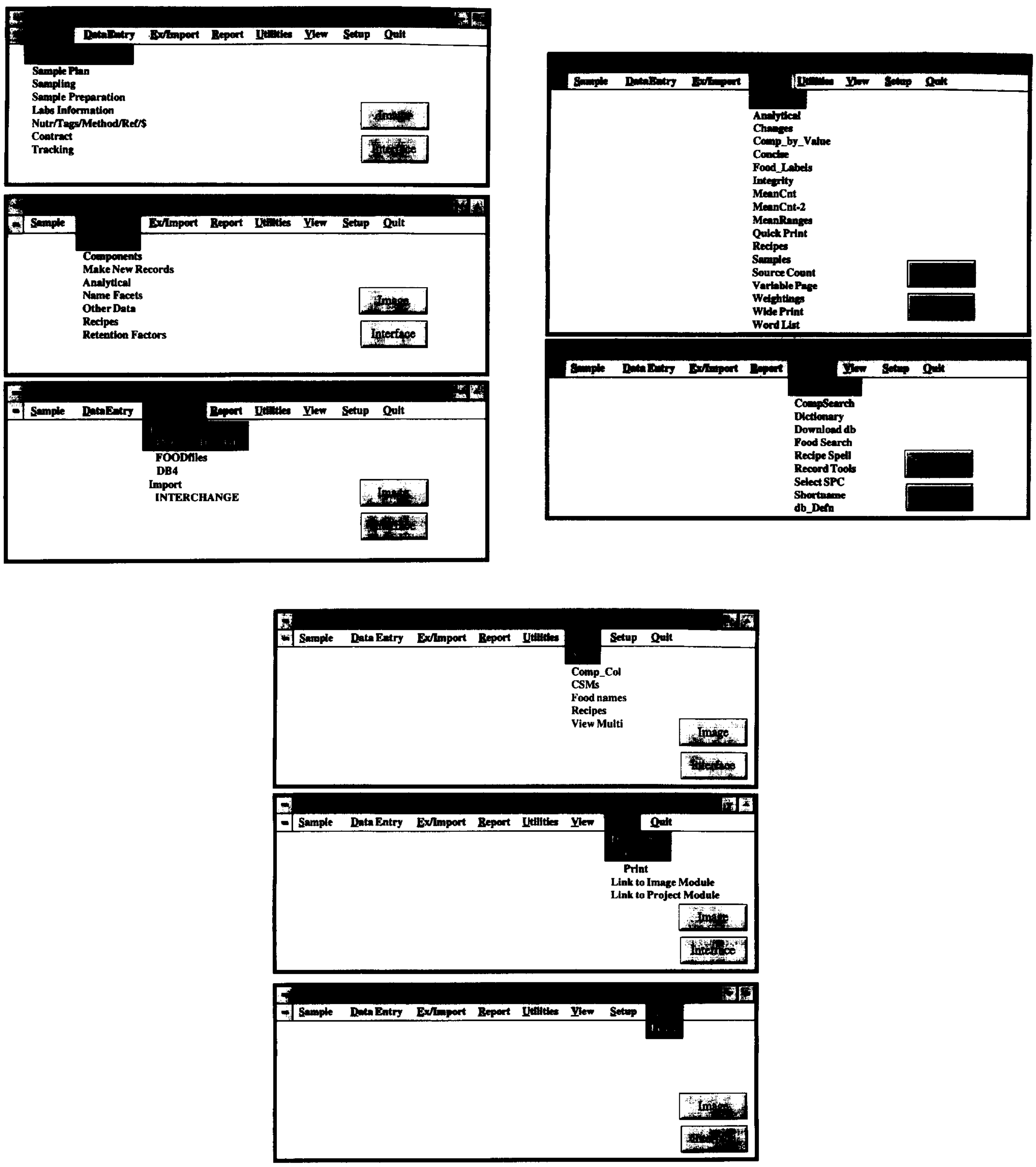
real change in a value that may be time-dependent in nature (e.g. a change in the food legislation specifying a different fat content for low-fat milk, effective on a particular date).

In addition to a source code for each value, which identifies the origin of the data, data quality indicators have been added. At the moment, these are alphabetical character codes that identify the data type (analytical and several types of imputed) and express a few levels of confidence in the presented information. These will continue to be used until guidelines are developed by the IUNS/INFOODS Data Quality Expert Committee. Easy conversion will be possible once international standards are in place.

In some countries there are separate databases for 'reference' and for 'survey' (Perloff, 1990). In New Zealand there is one database with several well-defined levels of data aggregation, and source/confidence codes to differentiate analysed values from imputed values.

Compact disks have become an important part of our information systems. For example, the records for the CHANGES program, all new entries as well as all

Table 4.



changes in existing records, are so large it has been written to CD-ROM. Our images library, actual food samples that are photographed and digitized, is also stored on and used from a CD (Burlingame & Cook, 1994; Burlingame *et al.*, 1995*a*). The CD drive interacts easily with the information systems contained on a file server. Some institutional users of the food composition data have been issued with image CDs, where the same key is assigned to the data files, the descriptor files and the image files.

Table 4 shows the drop-down menus used in the system.

BOOKS AND COMPUTER PRODUCTS

Making food composition data widely available is the most important 'service' aspect of this work, but is also essential as the 'technology/information transfer' of our research activities. Data dissemination from this database takes several different forms, designed to meet the needs of a variety of users. Because of the versatility of the database and the common subroutines in the programming, each standard format can be customized for foods, nutrients and documentation details, and codes can easily be written for completely new formats (both electronic and print). Table 3 shows the five standard outputs, each of which can be produced with a few keystrokes.

FOODfiles (Burlingame *et al.*, 1995*b*) is most frequently purchased as a site licence, and is used by approximately 25 organizations including universities and large hospitals, and by commercial software developers. The set of seven relational files that make up FOODfiles is usually integrated into an existing database management system or is used to develop specialist systems for users.

Several companies have licensing arrangements with Crop and Food Research to use FOODfiles in their products. Diet1/NZ is used in several university departments, in the dietetics departments of more than one-half of the New Zealand hospitals, and by individuals in the food industry, health care sector and public relations companies (Table 4).

Of the three printed formats, *The Concise Tables* (Burlingame *et al.*, 1994) is the most popular. In New Zealand, virtually every dietitian, nutritionist, food scientist and tertiary student in these areas owns a copy.

CONCLUSIONS

Food composition databases are necessary for a variety of purposes including food trade, agriculture policy development, clinical care and research, epidemiological and experimental research, public health research and policy development, food product development, food service management and much more. In order to maximize the usefulness of a food database beyond the narrow use in dietary assessment, a multidisciplinary team approach is needed to keep up with advancements in all areas that impinge upon this kind of work. Additionally, the internationalization of research, policy development, food trade, etc., demands a certain amount of standardization; hence, the role of INFOODS is fundamental.

REFERENCES

Burlingame, B. A. (1991*a*). Using International Nutrient Data. In *Proceedings of the Sixteenth Nutrient Databank Conference, Nutrient Databases for the 1990s*, pp. 143–4.

Burlingame, B. A. (1991*b*). Food analysis and the development of the New Zealand food composition databases. *Proc. Nutr. Soc. Aust.*, **16**, 104–107.

Burlingame, B. A. (1993). International Interchange. In *Proceedings of the third OCEANIAFOODS Conference, December 1991*, pp. 117–21.

Burlingame, B. A. & Cook, F. M. (1994). Images in data bases. In *Proceedings of the 19th National Nutrient Databank Conference*, pp. 45–9.

Burlingame, B., Cook, F. & Duxfield, G. (1990). New Zealand's food composition databases. *Proc. Nutr. Soc. NZ*, **15**, 132–143.

Burlingame, B. A., Cook, F. M., Duxfield, G. M. & Milligan, G. C. (1995*a*). Food data: numbers, words and images. In *Quality and Accessibility of Food-related Data*, Vol. I, ed. H. Greenfield. AOAC International, Arlington, pp. 175–82.

Burlingame, B. A., Milligan, G. C., Apimerika, D. E. & Arthur, J. M. (1994). *The Concise New Zealand Food Composition Tables*, 2nd edn. New Zealand Institute for Crop and Food Research Ltd.

Burlingame, B. A., Milligan, G. C., Quigley, R. J. & Spriggs, T. (1995*b*). *FOODfiles Manual.* New Zealand Institute for Crop and Food Research Ltd.

Greenfield, H. & Southgate, D. A. T. (1992). *Food Composition Data: Production, Management and Use.* Elsevier Science, Barking, UK.

Klensin, J. C. (1992). *INFOODS: Food Composition Data Interchange Handbook.* United Nations University Press, Tokyo.

Klensin, J. C., Feskanich, D., Lin, V., Truswell, A. S. & Southgate, D. A. T. (1989). *Identification of Food Components for Data Interchange.* United Nations University, Tokyo.

Lewis, J. (1996). Redevelopment of the Australian Nutrient Databank. In *Proceedings of the fourth OCEANIAFOODS Conference*, Suva, Fiji, pp. 183–186.

Milligan, G. C. & Burlingame, B. A. (1991). Creation of the New Zealand food composition interim database. *Proc. Nutr. Soc. NZ*, **16**, 181–188.

Perloff, B. (1990). USDA's National Nutrient Data Bank. In *Proceedings of the Fifteenth National Nutrient Databank Conference*, pp. 11–17.

Petersen, B. (1994). Retrieving food data using the International Interface Standard for food databases. In *Report of the Third Annual Meeting of the FLAIR Eurofoods—Enfant Project, Vilamora, Portugal, 10–12 November 1993*, eds J. Castenmiller & C. E. West. Wageningen, pp. 170–172.

Rand, W. M., Pennington, J. A. T., Murphy, S. P. & Klensin, J. C. (1991). *Compiling Data for Food Composition Data Bases.* United Nations University, Tokyo.

Shields, H. M., Burlingame, B. A. & Dunne, P. (1988). The launching of the New Zealand Food Composition Tables and Database. *Proc. Nutr. Soc. NZ*, **13**, 160–164.

Truswell, A. S., Bateson, D. J., Madafiglio, K. C., Pennington, J. A. T., Rand, W. M. & Klensin, J. C. (1991). INFOODS Guidelines for Describing Foods: a systematic approach to describing foods to facilitate international exchange of food composition data. *J. Food Composit. Anal.*, **4**(1), 18–38.